Red meat intake interacts with a TGF-β-pathway-based polygenic risk score to impact colorectal cancer risk: Application of a novel approach for polygenic risk score construction.

Author Block:

Joel Sanchez Mendez[1], Bryan Queme[2], Yubo Fu[2], John Morrison[2], Juan P. Lewinger[2], Eric Kawaguchi[2], Huaiyu Mi[2], Mireia Obón-Santacana[3,4,5], Ferran Moratalla-Navarro[3,4,5,6],Vicente Martín[5,7,8], Victor Moreno[3,4,5,6], Yi Lin[9], Stephanie A Bien[9], Conghui Qu[9], Yu-Ru Su[9], Emily White[9,10], Tabitha A Harrison[9], Jeroen R Huyghe[9], Catherine M Tangen[9], Polly A Newcomb[9,10], Amanda I Phipps[9,10], Claire E Thomas[9], David V Conti[2], Jun Wang[2], Elizabeth A Platz[11], Temitope O Keku[12], Christina C. Newton[13], Caroline Y Um[13], Anshul Kundaje[14,15], Anna Shcherbina[14,15], Neil Murphy[16], Marc J Gunter[16,17], Niki Dimou[16], Nikos Papadimitriou[16], Stéphane Bézieau[18], Franzel JB van Duijnhoven[19], Satu Männistö[20], Gad Rennert[21,22,23], Alicja Wolk[24], Michael Hoffmeister[25], Hermann Brenner[25,26,27], Jenny Chang-Claude [28,29], Yu Tian[28,30], Loïc Le Marchand[31], Michelle Cotterchio[32], Konstantinos K. Tsilidis[33], D Timothy Bishop[34], Yohannes Adama Melaku[35,36], Brigid M. Lynch[36,37], Daniel D Buchanan[38,39,40], Cornelia M. Ulrich[41,42], Jennifer Ose[41,42], Anita R. Peoples[41,42], Andrew J Pellatt[43], Li Li[44], Matthew AM Devall[44], Peter T Campbell[45], Demetrius Albanes[46], Stephanie J Weinstein[46], Sonja I Berndt[46], Stephen B Gruber[47], Edward Ruiz-Narvaez[48], Mingyang Song[49,50], Amit D Joshi[49,50], David A Drew[50], Jessica L Petrick[51], Andrew T Chan[49,50,52,53,54,55], Marios Giannakis[54,56], Li Hsu[9,57], Ulrike Peters[9,10], W. James Gauderman[2], Mariana C. Stern[1].

[1]Department of Population and Public Health Sciences & USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[2]Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, California, USA.
[3]Unit of Biomarkers and Susceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), L'Hospitalet del Llobregat, 08908 Barcelona, Spain.
[4]ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908 Barcelona Spain.
[5]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain.
[6]Department of Clinical Sciences, Faculty of Medicine and health Sciences and Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona (UB), L'Hospitalet de Llobregat, 08908 Barcelona, Spain.
[7]The Research Group in Gene – Environment and Health Interactions (GIIGAS) / Institut of Biomedicine (IBIOMED), Universidad de León, 24071 León, Spain.
[8]Faculty of Health Sciences, Department of Biomedical Sciences, Area of Preventive Medicine and Public Health, Universidad de León, 24071 León, Spain.
[9]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA.
[10]Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA.
[11]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.
[12]Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina, USA.
[13]Department of Population Science, American Cancer Society, Atlanta, Georgia.
[14]Department of Genetics, Stanford University, Stanford, California, USA.
[15]Department of Computer Science, Stanford University, Stanford, California, USA.
[16]Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France.
[17]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.
[18]Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France.
[19]Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands.
[20]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland.
[21]Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel.
[22]Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel.
[23]Clalit National Cancer Control Center, Haifa, Israel.
[24]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

[25]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[26]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.
[27]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.
[28]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
[29]University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany.
[30]School of Public Health, Capital Medical University, Beijing, China.
[31]University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
[32]Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada
[33]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.
[34]Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK.
[35]Flinders Health and Medical Research Institute, Adelaide Institute for Sleep Health, Flinders University, Adelaide, South Australia, Australia.
[36]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia.
[37]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia.
[38]Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria 3010 Australia.
[39]University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria 3010 Australia.
[40]Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, Victoria, Australia.
[41]Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA.
[42] Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA.
[43]Department of Cancer Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas, USA.
[44]Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA.
[45]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA.
[46]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA.
[47]Department of Medical Oncology & Therapeutics Research, City of Hope National Medical Center, Duarte CA, USA.
[48]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.
[49]Departments of Epidemiology and Nutrition, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.
[50]Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

[51]Slone Epidemiology Center at, Boston University, Boston, Massachusetts, USA.
[52]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[53]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[54]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
[55]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA.
[56]Dana-Farber Cancer Institute, Harvard Medical School, Harvard University, Boston, Massachusetts, USA.
[57]Department of Biostatistics, University of Washington, Seattle, Washington, USA.

Running title: Interactions of colorectal cancer pathway-based polygenic risk scores with meat intake.

Corresponding authors: Mariana C. Stern, marianas@usc.edu; W. James Gauderman, jimg@usc.edu.

Correspondence: Mariana C. Stern, University of Southern California, Keck School of Medicine, Norris Comprehensive Cancer Center, 1441 Eastlake Avenue, room 4455, Los Angeles, CA 90089. marianas@usc.edu, Tel: +1 (323) 865 0811. W. James Gauderman, Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California, 1845 N Soto St, room SSB 202K, Los Angeles, CA 90032. jimg@usc.edu, Tel: + 1 (323) 442 1567.

Conflicts of Interest:

C.M.U. has as cancer center director oversight over research funded by several pharmaceutical companies but has not received funding directly herself.

WORD COUNT: 4560 (Inclusive of Materials AND Methods section)

ABSTRACT:

Background:

High intake of red and/or processed meat are established colorectal cancer (CRC) risk factors. Genome-wide association studies (GWAS) have reported 204 variants (G) associated with CRC risk. We used functional annotation data to identify subsets of variants within known pathways and constructed pathway-based Polygenic Risk Scores (pPRS) to model pPRS x environment (E) interactions.

Methods:

A pooled sample of 30,812 cases and 40,504 CRC controls of European ancestry from 27 studies were analyzed. Quantiles for red and processed meat intake were constructed. The 204 GWAS variants were annotated to genes with AnnoQ and assessed for overrepresentation in PANTHER-reported pathways. pPRS's were constructed from significantly overrepresented pathways. Covariate-adjusted logistic regression models evaluated pPRSxE interactions with red or processed meat intake in relation to CRC risk.

Results:

A total of 30 variants were overrepresented in four pathways: Alzheimer disease-presenilin, Cadherin/WNT-signaling, Gonadotropin-releasing hormone receptor, and TGF-β signaling. We found a significant interaction between TGF-β-pPRS and red meat intake ($p = 0.003$). When variants in the TGF-β pathway were assessed, significant interactions with red meat for rs2337113 (intron *SMAD7* gene, Chr18), and rs2208603 (intergenic region *BMP5*, Chr6) ($p = 0.013$ & 0.011, respectively) were observed. We did not find evidence of pPRS x red meat interactions for other pathways or with processed meat.

Conclusions:

This pathway-based interaction analysis revealed a significant interaction between variants in the TGF-β pathway and red meat consumption that impacts CRC risk.

Impact:

These findings shed light into the possible mechanistic link between CRC risk and red meat consumption.

Impact statement:

In this work, we developed pathway-based Polygenic Risk Scores, which for the first time suggested, that red meat intake interacts with variants overrepresented in TGF-β signaling pathway to increase colorectal cancer risk.

INTRODUCTION:

Colorectal Cancer (CRC) has the third highest cancer incidence, and the second highest cancer mortality worldwide (1). It is estimated that if current trends are maintained, 3.2 million new cases of CRC will occur in 2040, with a concomitant 73.4% increase in deaths to over 1.6 million per annum (2). An ever-expanding body of evidence has identified several modifiable/lifestyle risk factors for CRC. Examples that have been associated with differential CRC risk include (but are not limited to) smoking, alcohol consumption, decreased dietary consumption of fiber, whole grains, vegetables, and fruits, increased intake of red meat and processed meat, living with obesity and overweight, and inflammatory bowel disease (3–11). Furthermore, red meat and processed meat have been labeled as “probably carcinogenic to humans” (group 2a carcinogen), and “carcinogenic to humans” (group 1 carcinogen) by the International Agency for Research on Cancer (IARC) due to the mounting evidence vis-à-vis CRC (12). Different putative causal pathways have been proposed for the association between red meat and/or processed meat with CRC. Mainly, the presence of heme iron combined with long-term intestinal dysbiosis, in addition to the neoformation of mutagenic compounds (N-nitroso compounds, heterocyclic amines, polycyclic aromatic hydrocarbons) have been reported to act both systemically, and locally at the mucosa of the colon and rectum (12–22); thus, potentially mediating the association between red meat and/or processed meat consumption and CRC risk.

Albeit the majority of CRC cases are considered sporadic, three-quarters of the patients have a negative family history (23). Genome-wide association studies (GWAS) have informed a heritable basis for the disease(24), with 205 single-nucleotide polymorphisms (SNPs) showing a marginally statistically significant association with the trait (25). Moreover, non-additive estimates of heritability such as gene-environment (GxE) interactions have been reported to

explain some variance in the heritability of CRC (24,26). GxE interactions refers to the scenario where the effect of a genetic variant differs across levels of the environmental exposure, or vice versa. That the susceptibility of a person to a given environmental exposure can be influenced by a genetic variant. Previously, most studies that evaluated GxE in CRC were considered underpowered to detect GxE interactions (27). However, a recently published genome-wide GxE interaction analysis reported two significant interactions for red meat consumption: the rs4871179 SNP in chromosome (chr) 8 (downstream of *HAS2*) and the rs35352860 SNP in chr18 (*SMAD7* intronic region) (28). These genome-wide GxE interactions could potentially explain the increased risk of CRC in certain population subgroups. However, how to adequately identify individuals with an elevated CRC risk, and the underlying mechanistic explanation for this highly polygenic carcinogenic process is not readily apparent through traditional candidate GxE analyses that sequentially evaluate variants associated with the trait.

Polygenic Risk Scores (PRS) permits to aggregate multiple SNPs reported through GWAS with a typically small effect to estimate heritability or genetic liability to a particular phenotype. In a classic PRS, a sum of the risk alleles that an individual carries, weighted by the effect sizes of the variants (e.g., GWAS summary statistics), is computed (29). This enables output of a risk estimate at the individual level that could be leveraged to stratify patients and provide targeted screening, with the potential to guide precise CRC prevention (30–32). However, when modeled through association analyses in relation to a phenotype, this canonical PRS approach assumes that every SNP that is an element of the PRS (i.e., $SNP_i \in PRS$), exhibits the same main effect size (i.e., $\beta_{SNP_i} = \cdots = \beta_{SNP_n}$). Moreover, a loss of information ensues, as the PRS groups SNPs that are associated to the phenotype via dissimilar biological processes. To address this challenge, pathway-based PRSs constructs gene sets that aggregate risk alleles across *a priori* defined pathways via existing annotated databases (e.g., REACTOME, PANTHER Pathway, KEGG (33–35)) (36). This strategy provides functionally informed risk scores that have enabled

detection of putative disrupted biological pathways for various polygenic traits (e.g., Alzheimer's disease, schizophrenia, coronary artery disease (37–39)) but has yet to be deployed for CRC risk.

In this study, we applied functional annotation data to identify subsets of GWAS variants that are components of biological pathways to construct pathway-based PRS (pPRS). Additionally, we evaluated multiplicative interactions between pPRS and red meat or processed meat intake and CRC in a large, pooled dataset to ascertain whether genes within specific pathways that exhibit a statistically significant interaction with the environmental factor could influence CRC risk.

MATERIALS and METHODS:

Study participants

Details on study participants, genotyping information, and environmental exposure data have been previously published (40–42). In brief, 27 studies from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Colorectal Cancer Transdisciplinary Study (CORECT), and the Colon Cancer Family Registry (CCFR, (43)) were utilized. These studies have been previously described (28,40). The outcome/case status was defined as a positive diagnosis of invasive colorectal adenocarcinoma (International Classification of Diseases codes (ICD)153-154), and confirmed via medical records, pathology reports, or death certificate information. Nested case-control sets were assembled via risk-set sampling in cohort studies, and population-based controls were matched on age, sex, race, and enrollment date/trial group (in the SELECT trial) in case-control studies. Cases were further stratified by tumor location as follows: proximal colon (ICD-9 codes: 153.0, 153.1, 153.4, and 153.6), distal colon (ICD-9 codes: 153.2, 153.3, 153.7), and rectum (ICD-9 codes: 154.0, 154.1). The study sample was restricted to only include individuals of European ancestry according to self-reported race and

ethnicity. Exclusion criteria included: duplicate participants or cryptic relatedness, errors on genotyping and/or imputation, age outliers, diagnosis of advanced adenoma (only), and missing data for red meat and processed meat intake. All participants gave written informed consent, and studies were approved by their individual institutional review boards.

Genotype data

Genotyping arrays utilized by each study are described in Supplementary Table 1. Quality control and genotyping metrics were conducted as previously described (41,42). Briefly, samples were excluded with the following criteria: 1) Single Nucleotide polymorphisms (SNPs) with missing call rate >2-5%; 2) Departure from Hardy-Weinberg equilibrium (threshold = P<1x10-4); 3) Mismatch between genotyped and reported sex; 4) Discordant genotype calls with duplicate samples. The Haplotype Reference Consortium panel of 39.1 million variants was used to impute to the genotypes with the University of Michigan Imputation Server (44). We used the Binarydosage R package (https://cran.r-project.org/web/packages/BinaryDosage) to process and manage the imputed genotypes in our sample. Imputed SNPs with a minor allele frequency (MAF) <1% and/or imputation accuracy of $R^2 \leq 0.8$ were dropped. PLINK 1.9 was utilized to compute principal component analysis on 30,000 randomly sampled imputed SNPs with MAF >5% and $R^2 > 0.99$ to control for population stratification (45). For use in polygenic risk score analyses, we leveraged data from 204 autosomal SNPs that have been previously reported to showcase a marginal genome-wide statistically significant association with CRC (P<5x10-8) (25).

Environmental exposure data

In-person interviews, phone interviews, or structured self-administered questionnaires were utilized to collect environmental risk factors and sociodemographic variables as previously described (40). After quality-control checks in each study were conducted, common data

elements (CDEs) were defined (46). Study-specific data elements were mapped to the CDEs through an iterative process with outreach to data contributors. Via SAS and a T-SQL script, all CDE were pooled into a dataset that included common definitions, standardized permissible values, and standardized coding. The dataset was evaluated for quality assurance, with the distribution of the data within and between studies assessed for outlier observations. Food frequency questionnaires were utilized to compute total energy intake (kcal/day). Mean imputation was utilized for studies with partial missingness. In studies that did not report the variable, the value was set to zero. Body mass index (BMI) was computed as weight (kg)/height(cm)^2. The predefined World Health Organization BMI cut-off points were used to generate an ordinal categorical variable: normal weight (18.5 to <25 kg/m^2), overweight (≥25.0 to <30 kg/m^2), and obese (≥30 kg/m^2) (47). For the main exposure of interest, a pooled variable to track red meat intake and processed meat intake was developed. The first was defined as consumption of beef, pork, and/or lamb, while the latter included intake of bacon, sausages, luncheon/deli meats, and/or hot dogs. Both variables were expressed as servings per day (i.e., servings/day), with one serving equivalent to 70.9 grams or 2.5 ounces. Heterogeneity in the red meat/processed meat definition for the individual studies caused overlap, with some studies including processed meat under the red meat definition. Red/processed meat variables were further operationalized by constructing sex-study-specific quartiles, with the median of intake for each quartile utilized to define consumption values, fitted as continuous predictors in regression models.

Pathway overrepresentation

From the 204 genome-wide significant autosomal SNPs reported by Fernandez-Rozadilla et.al (25), rsIDs were extracted in accordance with the genome reference consortium (GRC) human build 37 (GRCh37). These SNPs were then annotated using the Annotation Query

(AnnoQ) platform, which provides integrated functional annotation supported by a database of ~39 million pre-annotated human variants from the Haplotype Reference Consortium (48). Annotations to Ensembl and Reference sequence (RefSeq) genes were derived using inferences from Annotate Variation (ANNOVAR), SnpEff, and Ensembl Variant Effect Predictor (VEP) (49–53). SNPs residing in enhancer regions were linked to their target genes via the Predicted by Experimental Results: Enhancer-Gene Relationships Illustrated by a Nexus of Evidence (PEREGRINE), which leverages experimental data from chromatin interaction analysis by paired-end tag sequencing, expression quantitative trait locus, and Hi-C across 78 cell and tissue types (54). After duplicates were removed, variants were uploaded into the Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification system v.18.0. (55,56). which is built on 143 annotated complete genomes and assigns sequences to PANTHER Pathways (v3.6.7; 177 pathways and 3092 components) (34). The set of genes falling within each pathway was tested for overrepresentation relative to the PANTHER Pathway annotation sets using Fisher's exact test ($p < 0.05$), with false discovery rate correction applied for multiple testing (57). Each significantly overrepresented pathway was subsequently used for downstream analyses.

Statistical analyses

Standard polygenic risk score (PRS) methods (29) were used to compute pathway-based polygenic risk scores (pPRS). SNP-specific weights were extracted after regressing CRC on the set of 204 variants in our sample with a logistic regression model that adjusted for study, sex, age, and three principal components of ancestry. A complete list of variants and weights can be found in Supplementary Table 2. The subset of weights corresponding to each of the

overrepresented pathways were used to compute the pPRS. To identify pPRS by red meat/processed meat intake interactions (pPRSxE), a 1 degree-of-freedom (1-df) test of gene-environment (GxE) interaction was utilized in multivariable logistic regression. All models were adjusted for study, age, sex, and the first three principal components to account for population structure. Simulations have shown that Type I error rates are conserved when the same sample is used to both generate PRS weights and to evaluate potential pPRSxE interactions (58). As variants can simultaneously participate in multiple pathways, pPRS subsets that included unique variants, and variants that showed overlap to other pathways were generated. In addition, we generated a PRS that included the 174 variants that were not overrepresented among PANTHER-defined pathways. Potential effect modification of the association between red meat/processed meat intake and CRC risk by pPRS was evaluated by further stratifying the models by quartiles of the corresponding pPRS. All analyses were performed using the R programming language (v.4.2.1). All p-values are two-sided, with a value < 0.05 deemed statistically significant.

Data availability

Data generated are available upon request from the corresponding author.

RESULTS:

A total of 30,812 cases and 40,504 controls met the inclusion criteria. Of these, 138 had missing information on red meat intake (cases = 69, controls = 69) and 5,764 had missing processed meat intake information (cases = 2,634, controls = 3,130). Tumors were more commonly located in the proximal colon (35.2%), followed by the distal colon (30.4%), and the rectum (27.9%); 6.5% of the cases had a missing tumor location. Compared to controls, CRC cases were on average older, had a higher mean BMI, were more likely to be male, and reported a higher total

caloric intake. Cases had a higher mean intake of red meat (0.603 servings/day ± 0.463), and processed meat (0.350 servings/day ± 0.344) compared to controls (0.537 servings/day ± 0.437 & 0.276 ± 0.286, respectively, P-values < 0.001). Participant characteristics are described in detail in Supplementary Table 3.

Pathway overrepresentation analysis and pathway-based polygenic risk scores

Pathway overrepresentation analysis revealed that 30 SNPs were statistically significantly overrepresented across four biological pathways: 1) Gonadotropin-releasing hormone receptor pathway, n = 16 SNPs; 2) Alzheimer disease-presenilin pathway, n = 8 SNPs; 3) Cadherin/WNT signaling pathway, n = 7 SNPs; and 4) Transforming Growth Factor – beta (TGF-β) signaling pathway, n = 14 SNPs (Supplementary Table 4). Variants can simultaneously be overrepresented in multiple pathways (Figure 1). The biggest overlap was reported between SNPs in the Gonadotropin-releasing hormone receptor pathway and TGF-β signaling pathway (n = 9). The rs35470271 SNP, which maps to chromosome 3 in the intergenic region of the *ZNF621* (Zinc Finger Protein 621), was the most commonly overrepresented SNP, with involvement in all pathways except TGF-β signaling (Figure 1).

Overrepresented and uniquely overrepresented SNPs in the four biological pathways were leveraged to construct pPRS. These pPRS's were contrasted with a general PRS which contained all 204 variants that have been *a priori* marginally associated with CRC risk (overall PRS). As expected, moderate to strong correlations were found between pPRS's and their corresponding pPRS that contained only uniquely overrepresented SNPs (Figure 2). A strong correlation was found between the Alzheimer disease-presenilin-pPRS and Cadherin/WNT signaling-pPRS (r^2 = 0.69), and between the TGF-β signaling-pPRS (TGF-β-PRS) and

Gonadotropin-releasing hormone receptor-pPRS (r^2 = 0.62) (Figure 2), driven by variants that overlapped across said pathways (Figure 1).

Interaction between pPRS with red meat or processed meat intake in CRC risk

Evaluation of the putative multiplicative interaction between pPRS's with red meat or processed meat intake (pPRSxE) in relation to CRC risk, revealed a statistically significant interaction for the overall PRS with red meat intake (OR = 0.95; 95% CI = 0.91-0.98; p-value = 0.0048), and a statistically significant interaction for the TGF-β-PRS with red meat intake (OR = 0.95; 95% CI = 0.92-0.98; p-value = 0.0031). Moreover, the interaction remained significant when assessed in the TGF-β-PRS with only uniquely overrepresented variants for the TGF-β signaling pathway (OR = 0.94; 95% CI = 0.91-0.98; p-value = 0.0007). No statistically significant interactions were identified for processed meat intake (Figure 3).

When the pPRSxE interactions were examined in a multivariable multinomial logistic regression model with tumor location as the outcome variable (i.e., proximal colon, distal colon, and rectum), a statistically significant interaction was observed between the overall PRS, and the TGF-β-PRS and red meat intake in relation to CRC risk in the proximal colon and in the rectum (all p-values < 0.05) (Supplementary Table 5). In the TGF-β-PRS with only uniquely overrepresented variants for the TGF-β signaling pathway, a statistically significant interaction with red meat intake was identified for adenocarcinomas in the proximal and distal colon (p-value = 0.0152 & p-value = 0.0245, respectively). Furthermore, we observed a statistically significant interaction between the TGF-β-PRS and processed meat intake for proximal colon adenocarcinomas (p-value = 0.0488) (Supplementary Table 5).

We stratified the association between red meat/processed meat intake in relation to CRC risk by quartiles of the distribution of the pPRS's (Supplementary Table 6). After controlling for confounders, individuals in the ≤25th percentile of the overall PRS were 1.32 (95% CI = 1.21-1.44; p-value = 1.22x10-9), and 1.34 (95% CI = 1.16-1.55; p-value = 8.40x10-5) times as likely to have colorectal cancer associated with an increase in the median of servings/day from study-sex-specific quartiles of red meat and processed meat intake, respectively, whereas individuals in the ≥75th percentile of the overall PRS were 1.17 and 1.27 times as likely (red meat OR = 1.17; 95% CI = 1.07-1.28; p-value = 0.0005 & processed meat OR = 1.27; 95% CI = 1.10-1.47; p-value = 0.0011) (Table 1). Moreover, we observed that the effect of red meat intake was stronger for individuals with lower genetic risk according to the TGF-β-PRS distribution. Specifically, individuals in the first (Q1), second (Q2), third (Q3), and fourth quartile (Q4) of the TGF-β-PRS had a 37%, 33%, 27%, and 17% increase in CRC risk, respectively (Q1 OR = 1.37; 95% CI = 1.26-1.49; p-value = 1.20x10-13, Q2 OR = 1.33; 95% CI = 1.22-1.45; p-value = 4.39x10-11, Q3 OR = 1.27; 95% CI = 1.16-1.38; p-value = 5.45x10-08, Q4 OR = 1.17; 95% CI = 1.07-1.27; 4.41x10-04) (Table 1).

TGF-β-PRS and red meat intake interactions in CRC risk

We further interrogated the statistically significant interaction between the TGF-β-PRS and red meat intake to parse out the effect of the driver variants for this association. The interaction remained significant after the model was adjusted with a PRS constructed from variants that showed no overlap with the TGF-β signaling pathway (n = 16) (p-value = 0.0041), after the model was adjusted with a PRS for SNPs overrepresented in other pathways regardless of overlap with the TGF-β signaling pathway (n = 25) (p-value = 0.0039), and after the model was adjusted with a PRS constructed with variants that are not overrepresented in the TGF-β

signaling pathway, but are marginally associated with CRC risk (n = 190) (p-value = 0.003). Furthermore, when we fitted a PRS constructed with the variants in the intersection between the TGF-β signaling-pPRS and the Gonadotropin-releasing hormone receptor-pPRS (i.e., overlap between both pathways, n = 9), no evidence of an interaction with red meat intake was observed (p-value = 0.457) (Table 2). Sensitivity analyses confirmed that the TGF-β-PRS interaction with red meat was not confounded by smoking, BMI, or total caloric intake (kcal/day) (Supplementary Table 7).

When the variants that were uniquely overrepresented in the TGF-β signaling pathway were assessed, we found a statistically significant interaction with red meat consumption for rs2337113 which maps to chromosome 18 in an intronic region for the gene *SMAD7* (SMAD family member 7) (p-value = 0.0006), and for rs2208603, which maps to chromosome 6 in the intergenic region for *BMP5* (bone morphogenetic protein 5) (p-value = 0.036) (Table 3). Furthermore, we found no evidence of an interaction with red meat intake when the five uniquely overrepresented SNPs in TGF-β signaling pathway were removed from the overall PRS (OR = 0.96; 95% CI = 0.93-1.00; p-value = 0.0513).

DISCUSSION:

In this large pPRSxE analysis with more than 70,000 individuals of European ascend from 27 different studies, we report a statistically significant interaction between a PRS with red meat intake in relation to CRC risk. Furthermore, we constructed four pathway-based PRS's which when deployed, suggested for the first time, that the significance of the PRSxE interaction is specifically driven by variants overrepresented for the TGF-β signaling pathway (Figure 3).

The TGF-β signaling pathway plays a biological role in cell proliferation, differentiation, migration, and apoptosis (59). Many of these roles are context-dependent and paradoxical. In benign tissues, TGF-β signaling inhibits epithelial growth and cell proliferation, while promoting differentiation, and apoptosis (59,60). Classically associated with epithelial cancers, TGF-β signaling acts as a tumor suppressor via promotion of cell cycle arrest and apoptosis during the tumor initiation phase (60). However, it facilitates tumor cell proliferation, epithelial-mesenchymal transition, fibrosis, inflammation, and cellular stemness via activation of the canonical and noncanonical TGF-β signaling pathways during tumor progression (61,62). Moreover, autocrine TGF-β signaling activation has the ability to change the architecture of the tumor microenvironment and suppress anti-tumor immune-mediated responses, enabling cell invasion, dissemination, and therapeutic resistance (63,64).

In CRC, an extensive body of work has reported on the involvement of this signaling pathway at multiple stages of the natural history of disease. Elevated TGF-β levels in plasma and primary tumor tissue correlate with tumor stage and recurrence (65,66), while its inhibition can prevent metastasis by inducing a potent cytotoxic T-cell response in murine models (67). Moreover, when the consensus molecular subgroup (CMS) classification for CRC was proposed, a distinctive transcriptional signature for TGF-β signaling -which conveyed the worst relapse-free survival of all 4 CMS subgroups- was reported (i.e., CMS-4) (68,69). The potential for an interaction between the TGF-β signaling pathway with red meat intake in relation to CRC risk remains understudied. Nevertheless, it has been suggested that oxysterols and aldehydes, formed through the oxidation of saturated fat in red meat, may upregulate TGF-β1 expression in macrophages and fibroblasts in the tumor vicinity (70). This sustained exposure to elevated concentrations of TGF-β1 in the colonic mucosa could facilitate clonal proliferation. However, this theory is not specific to red meat intake, as other sources of saturated fat exist (71). Additionally, we found no evidence to suggest that the TGF- β- PRS x Red Meat interaction was

confounded by body mass index or total caloric intake (Supplementary Table 7). Therefore, it is unlikely that our findings are solely driven by increased saturated fat intake.

It is possible then, that the underlying mechanism by which the TGF-β signaling pathway predisposes to an increased CRC risk rests in mutations of downstream components (72). Moreover, elements of this pathway, including GREM1, SMAD3, SMAD7, SMAD9, BMP2, BMP4, and RHPN2 have been linked to CRC etiology (59). Thus, in our study, we propose that polymorphisms targeting key elements of the canonical TGF-β signaling pathway interact with red meat intake to increase CRC risk (Table 3).

The rs2337113 SNP, which maps to the *SMAD7* gene, codes for an intracellular protein that is considered a negative regulator of TGF-β1 (73). Located in a region (long arm) of chromosome 18 that is commonly deleted in CRC (74,75), multiple *SMAD7* polymorphisms have been associated with CRC risk in Asian and European populations (76–79). In addition, deletion of *SMAD7* was associated with a protective dose effect in overall survival and disease-free survival in 264 tumor biopsies of colorectal cancer patients, while amplification was associated with poor prognosis (80). The involvement of the rs2337113 variant in CRC risk is supported by the work of Tian et.al, (81). Furthermore, this consortium previously published a genome-wide statistically significant interaction between another *SMAD7* polymorphism (rs35352860, chr18:48927384) with red meat intake in relation to CRC risk, which showcases a strong correlation (ρ = 0.642; 95% CI = 0.638-0.646; p-value = 2.2E-16) with the rs2337113 SNP. In our previous publication, we hypothesized that inhibition of the liver-derived peptide hepcidin via overexpression of rs35352860 could prevent the internalization of the transmembrane receptor ferroportin with a concomitant increase in iron output into the bloodstream, which in turn, could facilitate free radical formation and a proinflammatory, procardiogenic state (28). This aligns with

the work by Stolfi et al., which demonstrated that SMAD7 silencing inhibited the growth of CRC cell lines both in vitro and in vivo after transplantation into immunodeficient mice. Additionally, a significant increase in SMAD7 expression was observed in 14 matched pairs of CRC and adjacent tissues (82).

The variant in the 6p12.1 region (rs2208603) resides in the intergenic region of the gene *BMP5,* which codes for a cytokine member of the bone morphogenetic proteins (BMPs)/growth differentiation factors; a protein class within the TGF-β superfamily (83). BMP5 binds to the activin receptor-like kinase (ALK) -2 and -6, which controls binding to BMP type II and I receptors to induce SMAD-dependent, or SMAD-independent, intracellular signaling (84). This signaling is inhibited by SMAD7, as it preferentially interacts with type I receptors (85). Albeit BMPs have been reported to exhibit cell-dependent and context-dependent inhibitory capabilities (86–88). In CRC, deep sequencing of sporadic CRC samples identified loss of *BMP5* as an early and CRC-specific event, while high expression correlated with longer survival. Moreover, knockdown of *BMP5* in the SW480 cell line (adenocarcinoma of the colon, Dukes' type B) promoted proliferation. Thus, suggesting its relevance as a tumor suppressor gene (89). However, mutational events are uncommon; only <4% prevalence of non-silent somatic mutations in CRC was found in TCGA samples (90). Further research is required to evaluate the potential interaction between *BMP5* and red meat intake in relation to CRC. But recent work from Xiao et al., highlighted the role of BMP5 as a regulator of hepcidin and systemic iron homeostasis. In BMP5-inactive murine models, exposure to an iron-poor or high-iron diet decreased hepcidin levels in young mice. This transcriptional inactivation of hepcidin was exacerbated when concomitant BMP6 inactivation ensued, leading to hepatic and extrahepatic iron overload. Notably, the iron overload became more severe in animals exposed to a high-iron diet (91). This further supports our *a priori* mentioned hypothesis in which functional inhibition of hepcidin could foster a pro-carcinogenic state, increasing CRC risk.

Although previous literature has published on the association between the Gonadotropin-releasing hormone pathway, Alzheimer disease-presenilin pathway, and Cadherin/WNT signaling pathway with CRC risk (26,68,92–94), we found no evidence of a statistically significant interaction with red meat or processed meat intake (Figure 3). However, the observed overlap of SNPs overrepresented across multiple pathways suggests crosstalk and co-regulation between biological pathways (Figure 1). Prime examples are the multilevel processes jointly regulated by the TGF-β signaling pathway and Cadherin/WNT signaling. In conjunction, these pathways control extracellular gradients of morphogens during embryonic development, regulate target gene expression at the nucleus, and protein-protein interactions in the cytoplasm. Moreover, TGF-β and WNT synergistically promote tumorigenesis (95). Thus, further research is required to elucidate the potential antagonistic or synergistic effects of multilevel, multifactor interactions across these biological pathways in relation to red meat and/or processed meat intake for CRC.

Our study has several strengths. We leveraged harmonized pooled data that underwent systematic quality control, enabling the consideration of tumor localization, study design, and potential confounders of the exposure and outcome in the putative causal pathway. Additionally, the implementation of the novel pathway-based PRS x E approach by Gauderman et.al (58) allowed us not only to detect an interaction between SNPs that previously showed a marginal association with CRC with red meat intake in a single score, but also to identify the variants driving this signal and underpin the pathway leading to increased CRC risk. This approach supports the construction of constrained PRSs with variants selected according to a priori established criteria (e.g., pathway overrepresentation) (Figure 2). This is in line with a previous study, that reported no linear correlation between the predictive accuracy of PRSs and the

number of SNPs in the score (96). We also recognize some limitations. Questionnaire data was utilized to construct the red and processed meat intake variables, and given that many of the included studies have a case-control design, participants were asked to report on intake 1-2 years before study selection. Thus, we cannot rule out the potential for misclassification bias of the exposure and/or recall bias. In cohort studies, exposures were assessed at the time of blood draw or buccal collection at study-specific times (97). Also, we utilize study-specific quartiles to evaluate meat intake, which does not account for absolute differences. However, this approach is a valid method in pooled analysis of nutritional exposures (98). In addition, the potential for unmeasured confounders derived from lifestyle and behavioral traits affecting our findings cannot be discarded. A key strength of our approach lies in its comprehensive nature: we integrate annotations from SnpEff, ANNOVAR, and VEP, drawing from both Ensembl and RefSeq databases to enhance the robustness of SNP-to-gene mapping. To further account for regulatory variants, we incorporated PEREGRINE-based enhancer–gene link annotations, enabling the capture of potential functional effects from non-coding SNPs. Although we recognize that incorporating study-specific eQTL or chromatin interaction data would offer an even finer resolution of regulatory effects, our multi-tool strategy substantially reduces annotation discrepancies and improves the accuracy of linking non-coding variants to their target genes. Nonetheless, a limitation of our approach is that alternative tools and databases, such as Reactome (reactome.org) or Gene Ontology (geneontology.org), offer different pathway definitions that may not fully overlap with those in PANTHER, albeit no gold standard exists. Finally, our findings may not be generalizable to other racial and ethnic populations besides those of European ancestry, which contributed to the individual studies.

CONCLUSION:

In summary, this pathway-based interaction analysis proposes, for the first time, that a statistical interaction between SNPs overrepresented in the TGF-β signaling pathway and red meat intake

impacts CRC risk. This finding provides a potential mechanistic explanation linking red meat consumption and CRC risk.

AKNOWLEDGMENTS:

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control participants, as well as all the physicians, technicians, and students. CCFR: The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies). CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study. Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services. CPS-II: The authors express sincere appreciation to all Cancer Prevention Study-II participants, and to each member of the study and biospecimen management group. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's Surveillance Epidemiology and End Results Program. The authors assume full responsibility for all analyses and interpretation of results. The views expressed here are those of the authors and do not necessarily represent the American Cancer Society or the American Cancer Society – Cancer Action Network. DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance. EDRN: We acknowledge all contributors to the development of the resource at University of Pittsburgh School of Medicine, Department of Gastroenterology, Department of Pathology, Hepatology and Nutrition and Biomedical Informatics. EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization. Harvard cohorts: The study protocol was approved by the institutional review boards of the Brigham and Women’s Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department of Medicine, Brigham, and Women's Hospital as home of the NHS. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention’s National Program of Cancer Registries (NPCR) and/or the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West

Virginia, Wyoming. The authors assume full responsibility for analyses and interpretation of these data. Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.
LCCS: We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study. NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff. PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible. Cancer incidence data have been provided by the District of Columbia Cancer Registry, Georgia Cancer Registry, Hawaii Cancer Registry, Minnesota Cancer Surveillance System, Missouri Cancer Registry, Nevada Central Cancer Registry, Pennsylvania Cancer Registry, Texas Cancer Registry, Virginia Cancer Registry, and Wisconsin Cancer Reporting System. All are supported in part by funds from the Center for Disease Control and Prevention, National Program for Central Registries, local states or by the National Cancer Institute, Surveillance, Epidemiology, and End Results program. The results reported here, and the conclusions derived are the sole responsibility of the authors. SELECT: We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT. WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

FUNDING:

MCS and JSM received support from awards U54CA233465 and U2CCA252971 from the National Cancer Institute. Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, U01 CA164930, R01 CA201407). This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704 and P30CA014089. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685. ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services. CLUE II funding was from the National Cancer Institute (U01 CA086308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG018033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. Maryland Cancer Registry (MCR) Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services. The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143237 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19

CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR. COLO2&3: National Institutes of Health (R01 CA060987). CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional review boards of Emory University, and those of participating registries as required. CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support. DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B). DALS: National Institutes of Health (R01 CA048998 to M. L. Slattery). EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and and Region Skåne and Region Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom). Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182).
Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726. LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167). MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital

status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. BMLynch was supported by MCRF18005 from the Victorian Cancer Agency. MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464). MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong. NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA066635 and P30 DK034987. PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA037429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute´s Distinguished Professor Award to Alicja Wolk. VITAL: National Institutes of Health (K05 CA154337). WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: https://www-whi-org.s3.us-west-2.amazonaws.com/wp-content/uploads/WHI-Investigator-Long-List.pdf

REFERENCES:

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

2. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023 Feb;72(2):338–44.

3. Petimar J, Smith-Warner SA, Rosner B, Chan AT, Giovannucci EL, Tabung FK. Adherence to The World Cancer Research Fund/American Institute for Cancer Research 2018 Recommendations for Cancer Prevention and Risk of Colorectal Cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2019 Sep;28(9):1469–79.

4. Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. J Nutr. 2020 Apr 1;150(4):663–71.

5. Key TJ. Fruit and vegetables and cancer risk. Br J Cancer. 2011 Jan 4;104(1):6–11.

6. van Duijnhoven FJB, Bueno-De-Mesquita HB, Ferrari P, Jenab M, Boshuizen HC, Ros MM, et al. Fruit, vegetables, and colorectal cancer risk: the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr. 2009 May;89(5):1441–52.

7. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of colorectal cancer risk factors. Cancer Causes Control CCC. 2013 Jun;24(6):1207–22.

8. Bai T, Peng J, Zhu X, Wu C. Vegetarian diets and the risk of gastrointestinal cancers: a meta-analysis of observational studies. Eur J Gastroenterol Hepatol. 2023 Nov 1;35(11):1244–52.

9. Ren HG, Luu HN, Liu Y, Wang DW, Guo X. High intake of cruciferous vegetables reduces the risk of gastrointestinal cancers: results from observational studies. Crit Rev Food Sci Nutr. 2023 Dec 5;1–7.

10. Hullings AG, Sinha R, Liao LM, Freedman ND, Graubard BI, Loftfield E. Whole grain and dietary fiber intake and risk of colorectal cancer in the NIH-AARP Diet and Health Study cohort. Am J Clin Nutr. 2020 Sep;112(3):603–12.

11. Park Y, Hunter DJ, Spiegelman D, Bergkvist L, Berrino F, van den Brandt PA, et al. Dietary fiber intake and risk of colorectal cancer: a pooled analysis of prospective cohort studies. JAMA. 2005 Dec 14;294(22):2849–57.

12. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi FE, Benbrahim-Tallaa L, et al. Carcinogenicity of consumption of red and processed meat. Lancet Oncol. 2015 Dec;16(16):1599–600.

13. Bastide NM, Pierre FHF, Corpet DE. Heme Iron from Meat and Risk of Colorectal Cancer: A Meta-analysis and a Review of the Mechanisms Involved. Cancer Prev Res (Phila Pa). 2011 Feb 1;4(2):177–84.

14. Zheng W, Lee SA. Well-done Meat Intake, Heterocyclic Amine Exposure, and Cancer Risk. Nutr Cancer. 2009;61(4):437–46.

15. Cheng T, Chaousis S, Kodagoda Gamage SM, Lam AKY, Gopalan V. Polycyclic Aromatic Hydrocarbons Detected in Processed Meats Cause Genetic Changes in Colorectal Cancers. Int J Mol Sci. 2021 Oct 11;22(20):10959.

16. Sumer G, Oz F. The Effect of Direct and Indirect Barbecue Cooking on Polycyclic Aromatic Hydrocarbon Formation and Beef Quality. Foods. 2023 Jan;12(7):1374.

17. Young VB, Schmidt TM. Overview of the Gastrointestinal Microbiota. In: Huffnagle GB, Noverr MC, editors. GI Microbiota and Regulation of the Immune System [Internet]. New York, NY: Springer New York; 2008 [cited 2023 Oct 17]. p. 29–40. (Advances in Experimental Medicine and Biology; vol. 635). Available from: http://link.springer.com/10.1007/978-0-387-09550-9_3

18. Bingham SA, Pignatelli B, Pollock JR, Ellul A, Malaveille C, Gross G, et al. Does increased endogenous formation of N-nitroso compounds in the human colon explain the association between red meat and colon cancer? Carcinogenesis. 1996 Mar;17(3):515–23.

19. Gurjao C, Zhong R, Haruki K, Li YY, Spurr LF, Lee-Six H, et al. Discovery and Features of an Alkylating Signature in Colorectal Cancer. Cancer Discov. 2021 Oct;11(10):2446–55.

20. Tappel A. Heme of consumed red meat can act as a catalyst of oxidative damage and could initiate colon, breast and prostate cancers, heart disease and other diseases. Med Hypotheses. 2007;68(3):562–4.

21. Cross AJ, Pollock JRA, Bingham SA. Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat. Cancer Res. 2003 May 15;63(10):2358–60.

22. Seiwert N, Adam J, Steinberg P, Wirtz S, Schwerdtle T, Adams-Quack P, et al. Chronic intestinal inflammation drives colorectal tumor formation triggered by dietary heme iron in vivo. Arch Toxicol. 2021 Jul;95(7):2507–22.

23. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. Nat Rev Dis Primer. 2015 Dec 17;1(1):15065.

24. Jiao S, Peters U, Berndt S, Brenner H, Butterbach K, Caan BJ, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet. 2014 Jul 15;23(14):3898–905.

25. Fernandez-Rozadilla C. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. Nat Genet. 2023;55.

26. Schubert SA, Morreau H, de Miranda NFCC, van Wezel T. The missing heritability of familial colorectal cancer. Mutagenesis. 2020 Jul 11;35(3):221–31.

27. Yang T, Li X, Montazeri Z, Little J, Farrington SM, Ioannidis JPA, et al. Gene-environment interactions and colorectal cancer risk: An umbrella review of systematic reviews and meta-analyses of observational studies. Int J Cancer. 2019 Nov 1;145(9):2315–29.

28. Stern MC, Sanchez Mendez J, Kim AE, Obón-Santacana M, Moratalla-Navarro F, Martín V, et al. Genome-Wide Gene–Environment Interaction Analyses to Understand the Relationship between Red Meat and Processed Meat Intake and Colorectal Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2024 Mar 1;33(3):400–10.

29. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep 1;15(9):2759–72.

30. Cross B, Turner R, Pirmohamed M. Polygenic risk scores: An overview from bench to bedside for personalised medicine. Front Genet. 2022 Nov 11;13:1000667.

31. Thomas M, Sakoda LC, Hoffmeister M, Rosenthal EA, Lee JK, Van Duijnhoven FJB, et al. Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. Am J Hum Genet. 2020 Sep;107(3):432–44.

32. Obón-Santacana M, Díez-Villanueva A, Alonso MH, Ibáñez-Sanz G, Guinó E, López A, et al. Polygenic risk score across distinct colorectal cancer screening outcomes: from premalignant polyps to colorectal cancer. BMC Med. 2021 Nov 8;19(1):261.

33. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2020 Jan 8;48(D1):D498–503.

34. Mi H, Thomas P. PANTHER Pathway: an ontology-based pathway database coupled with data analysis tools. Methods Mol Biol Clifton NJ. 2009;563:123–40.

35. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.

36. Choi SW, García-González J, Ruan Y, Wu HM, Porras C, Johnson J, et al. PRSet: Pathway-based polygenic risk score analyses and software. Cordell HJ, editor. PLOS Genet. 2023 Feb 7;19(2):e1010624.

37. Harrison JR, Foley SF, Baker E, Bracher-Smith M, Holmans P, Stergiakouli E, et al. Pathway-specific polygenic scores for Alzheimer's disease are associated with changes in brain structure in younger and older adults. Brain Commun. 2023 Oct 1;5(5):fcad229.

38. Pistis G, Vázquez-Bourgon J, Fournier M, Jenni R, Cleusix M, Papiol S, et al. Gene set enrichment analysis of pathophysiological pathways highlights oxidative stress in psychosis. Mol Psychiatry. 2022 Dec;27(12):5135–43.

39. Goodman MO, Cade BE, Shah NA, Huang T, Dashti HS, Saxena R, et al. Pathway-Specific Polygenic Risk Scores Identify Obstructive Sleep Apnea–Related Pathways Differentially Moderating Genetic Susceptibility to Coronary Artery Disease. Circ Genomic Precis Med. 2022 Oct;15(5):e003535.

40. Figueiredo JC, Hsu L, Hutter CM, Lin Y, Campbell PT, Baron JA, et al. Genome-Wide Diet-Gene Interaction Analyses for Risk of Colorectal Cancer. Amos CI, editor. PLoS Genet. 2014 Apr 17;10(4):e1004228.

41. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013 Apr;144(4):799-807.e24.

42. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019 Jan;51(1):76–87.

43. Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2007 Nov;16(11):2331–43.

44. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

46. Hutter CM, Chang-Claude J, Slattery ML, Pflugeisen BM, Lin Y, Duggan D, et al. Characterization of Gene–Environment Interactions for Colorectal Cancer Susceptibility Loci. Cancer Res. 2012 Apr 15;72(8):2036–44.

47. Doak CM, Wijnhoven TMA, Schokker DF, Visscher TLS, Seidell JC. Age standardization in mapping adult overweight and obesity trends in the WHO European Region. Obes Rev Off J Int Assoc Study Obes. 2012 Feb;13(2):174–91.

48. Liu Z, Mushayahama T, Queme B, Ebert D, Muruganujan A, Mills C, et al. Annotation Query (AnnoQ): an integrated and interactive platform for large-scale genetic variant annotation. Nucleic Acids Res. 2022 Jul 5;50(W1):W57–65.

49. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun 6;17(1):122.

50. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.

51. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep 1;38(16):e164.

52. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation | Nucleic Acids Research | Oxford Academic [Internet]. [cited 2025 May 27]. Available from: https://academic.oup.com/nar/article/44/D1/D733/2502674?login=false

53. Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, et al. Ensembl 2024. Nucleic Acids Res. 2023 Nov 11;52(D1):D891–9.

54. Mills C, Muruganujan A, Ebert D, Marconett CN, Lewinger JP, Thomas PD, et al. PEREGRINE: A genome-wide prediction of enhancer to gene relationships supported by experimental evidence. PLOS ONE. 2020 Dec 15;15(12):e0243791.

55. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc. 2019 Mar;14(3):703–21.

56. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 2022;31(1):8–22.

57. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc. 2019 Mar;14(3):703–21.

58. Gauderman WJ, Fu Y, Queme B, Kawaguchi E, Wang Y, Morrison J, et al. Pathway Polygenic Risk Scores (pPRS) for the Analysis of Gene-environment Interaction [Internet]. Genetics; 2024 [cited 2025 Feb 5]. Available from: http://biorxiv.org/lookup/doi/10.1101/2024.12.16.628610

59. David CJ, Massagué J. Contextual determinants of TGFβ action in development, immunity and cancer. Nat Rev Mol Cell Biol. 2018 Jul;19(7):419–35.

60. Itatani Y, Kawada K, Sakai Y. Transforming Growth Factor-β Signaling Pathway in Colorectal Cancer and Its Tumor Microenvironment. Int J Mol Sci. 2019 Nov 20;20(23):5822.

61. Luo J, Chen XQ, Li P. The Role of TGF-β and Its Receptors in Gastrointestinal Cancers. Transl Oncol. 2019 Mar;12(3):475–84.

62. Principe DR, Doll JA, Bauer J, Jung B, Munshi HG, Bartholin L, et al. TGF-□: Duality of Function Between Tumor Prevention and Carcinogenesis. JNCI J Natl Cancer Inst. 2014 Feb 7;106(2):djt369–djt369.

63. Derynck R, Turley SJ, Akhurst RJ. TGFβ biology in cancer progression and immunotherapy. Nat Rev Clin Oncol. 2021 Jan;18(1):9–34.

64. Elliott RL, Blobe GC. Role of transforming growth factor Beta in human cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2005 Mar 20;23(9):2078–93.

65. Friedman E, Gold LI, Klimstra D, Zeng ZS, Winawer S, Cohen A. High levels of transforming growth factor beta 1 correlate with disease progression in human colon cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 1995;4(5):549–54.

66. Tsushima H, Kawata S, Tamura S, Ito N, Shirai Y, Kiso S, et al. High levels of transforming growth factor beta 1 in patients with colorectal cancer: association with disease progression. Gastroenterology. 1996 Feb;110(2):375–82.

67. Tauriello DVF, Palomo-Ponce S, Stork D, Berenguer-Llergo A, Badia-Ramentol J, Iglesias M, et al. TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. Nature. 2018 Feb;554(7693):538–43.

68. Guinney J, Dienstmann R, Wang X, De Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015 Nov;21(11):1350–6.

69. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet. 2015 Apr;47(4):320–9.

70. Biasi F, Mascia C, Poli G. The contribution of animal fat oxidation products to colon carcinogenesis, through modulation of TGF-β1 signaling. Carcinogenesis. 2008 May;29(5):890–4.

71. Alisson-Silva F, Kawanishi K, Varki A. Human risk of diseases associated with red meat intake: Analysis of current theories and proposed role for metabolic incorporation of a non-human sialic acid. Mol Aspects Med. 2016 Oct;51:16–30.

72. Xu Y, Pasche B. TGF-β signaling alterations and susceptibility to colorectal cancer. Hum Mol Genet. 2007 Apr 15;16(R1):R14–20.

73. Troncone E, Marafini I, Stolfi C, Monteleone G. Transforming Growth Factor-β1/Smad7 in Intestinal Immunity, Inflammation, and Cancer. Front Immunol. 2018;9:1407.

74. Grady WM, Pritchard CC. Molecular alterations and biomarkers in colorectal cancer. Toxicol Pathol. 2014 Jan;42(1):124–39.

75. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990 Jun;61(5):759–67.

76. Huang Y, Wu W, Nie M, Li C, Wang L. SMAD7 polymorphisms and colorectal cancer risk: a meta-analysis of case-control studies. Oncotarget. 2016 Nov 15;7(46):75561–70.

77. Alidoust M, Hamzehzadeh L, Khorshid Shamshiri A, Afzaljavan F, Kerachian MA, Fanipakdel A, et al. Association of SMAD7 genetic markers and haplotypes with colorectal cancer risk. BMC Med Genomics. 2022 Jan 11;15(1):8.

78. Xiao Q, Chen J, Zhu J, Zeng S, Cai H, Zhu G. Association of several loci of SMAD7 with colorectal cancer: A meta-analysis based on case–control studies. Medicine (Baltimore). 2023 Jan 6;102(1):e32631.

79. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007 Nov;39(11):1315–7.

80. Boulay JL, Mild G, Lowy A, Reuter J, Lagrange M, Terracciano L, et al. SMAD7 is a prognostic marker in patients with colorectal cancer. Int J Cancer. 2003 Apr 20;104(4):446–9.

81. Tian J, Zhang M, Zhang F, Gao K, Lu Z, Cai Y, et al. Developing an optimal stratification model for colorectal cancer screening and reducing racial disparities in multi-center population-based studies. Genome Med. 2024 Jun 13;16(1):81.

82. Stolfi C, De Simone V, Colantoni A, Franzè E, Ribichini E, Fantini MC, et al. A functional role for Smad7 in sustaining colon cancer cell growth and survival. Cell Death Dis. 2014 Feb 20;5(2):e1073.

83. Structural studies of the TGF-βs and their receptors – insights into evolution of the TGF-β superfamily - ScienceDirect [Internet]. [cited 2024 Oct 23]. Available from: https://www-sciencedirect-com.libproxy1.usc.edu/science/article/pii/S0014579312004036

84. Katagiri T, Watabe T. Bone Morphogenetic Proteins. Cold Spring Harb Perspect Biol. 2016 Jun;8(6):a021899.

85. Ehata S, Yokoyama Y, Takahashi K, Miyazono K. Bi-directional roles of bone morphogenetic proteins in cancer: Another molecular Jekyll and Hyde? Pathol Int. 2013;63(6):287–96.

86. Ghosh-Choudhury N, Ghosh-Choudhury G, Celeste A, Ghosh PM, Moyer M, Abboud SL, et al. Bone morphogenetic protein-2 induces cyclin kinase inhibitor p21 and hypophosphorylation of retinoblastoma protein in estradiol-treated MCF-7 human breast cancer cells. Biochim Biophys Acta. 2000 Jul 21;1497(2):186–96.

87. Ghosh-Choudhury N, Woodruff K, Qi W, Celeste A, Abboud SL, Ghosh Choudhury G. Bone morphogenetic protein-2 blocks MDA MB 231 human breast cancer cell proliferation by inhibiting cyclin-dependent kinase-mediated retinoblastoma protein phosphorylation. Biochem Biophys Res Commun. 2000 Jun 16;272(3):705–11.

88. Miyazaki H, Watabe T, Kitamura T, Miyazono K. BMP signals inhibit proliferation and in vivo tumor growth of androgen-insensitive prostate carcinoma cells. Oncogene. 2004 Dec 16;23(58):9326–35.

89. Chen E, Yang F, He H, Li Q, Zhang W, Xing J, et al. Alteration of tumor suppressor BMP5 in sporadic colorectal cancer: a genomic and transcriptomic profiling based study. Mol Cancer. 2018 Dec 20;17(1):176.

90. Korkut A, Zaidi S, Kanchi RS, Rao S, Gough NR, Schultz A, et al. A Pan-Cancer Analysis Reveals High-Frequency Genetic Alterations in Mediators of Signaling by the TGF-β Superfamily. Cell Syst. 2018 Sep 26;7(4):422.

91. Xiao X, Xu Y, Moschetta GA, Yu Y, Fisher AL, Alfaro-Magallanes VM, et al. BMP5 contributes to hepcidin regulation and systemic iron homeostasis in mice. Blood. 2023 Oct 12;142(15):1312–22.

92. Aguilar-Rojas A, Huerta-Reyes M. Human gonadotropin-releasing hormone receptor-activated cellular functions and signaling pathways in extra-pituitary tissues and cancer cells (Review). Oncol Rep. 2009 Nov;22(5):981–90.

93. Ohlsson B. Gonadotropin-Releasing Hormone and Its Physiological and Pathophysiological Roles in Relation to the Structure and Function of the Gastrointestinal Tract. Eur Surg Res. 2016 Apr 19;57(1–2):22–33.

94. Ospina-Romero M, Glymour MM, Hayes-Larson E, Mayeda ER, Graff RE, Brenowitz WD, et al. Association Between Alzheimer Disease and Cancer With Evaluation of Study Biases: A Systematic Review and Meta-analysis. JAMA Netw Open. 2020 Nov 2;3(11):e2025515.

95. Guo X, Wang XF. Signaling cross-talk between TGF-β/BMP and other pathways. Cell Res. 2009 Jan;19(1):71–88.

96. Sassano M, Mariani M, Quaranta G, Pastorino R, Boccia S. Polygenic risk prediction models for colorectal cancer: a systematic review. BMC Cancer. 2022 Jan 15;22(1):65.

97. Gong J, Hutter CM, Newcomb PA, Ulrich CM, Bien SA, Campbell PT, et al. Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. Williams SM, editor. PLOS Genet. 2016 Oct 10;12(10):e1006296.

98. Kim DH, Smith-Warner SA, Spiegelman D, Yaun SS, Colditz GA, Freudenheim JL, et al. Pooled analyses of 13 prospective cohort studies on folate intake and colon cancer. Cancer Causes Control CCC. 2010 Nov;21(11):1919–30.

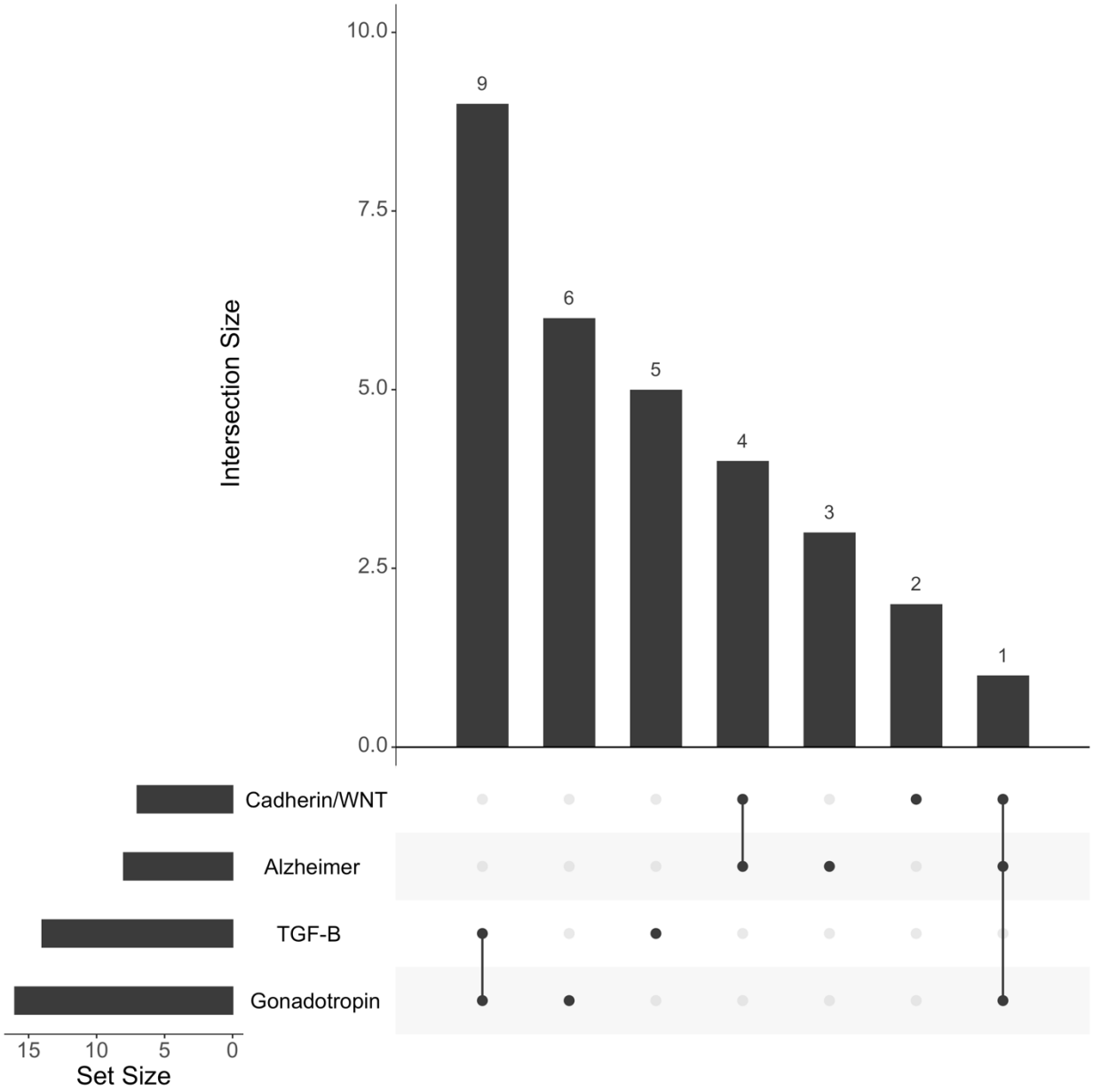

Figure 1. Pathways overrepresented in CRC, showing the number of variants associated with each pathway and the extent of overlap among annotated variants. CRC; colorectal cancer.
Note: Set size and intersection size correspond to the number of single nucleotide polymorphisms contained in each pathway-based Polygenic Risk Score (pPRS) and exhibiting overlap, respectively.

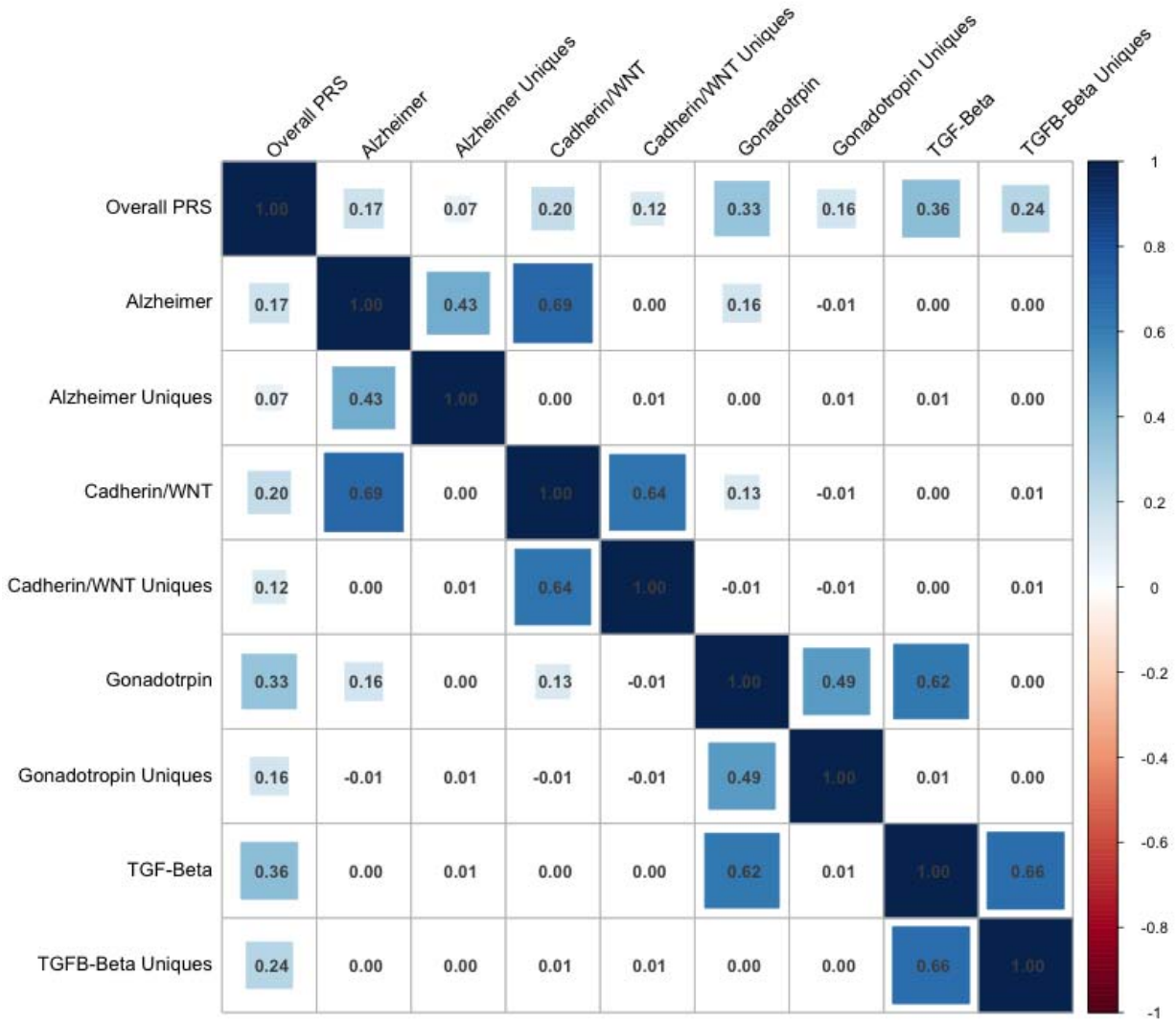

Figure 2. Correlogram for pathway-based polygenic risk scores (pPRS). The numbers correspond to Pearson's correlation coefficients.
Overall PRS contains 204 SNPs that have been previously associated with CRC risk.
CRC; colorectal cancer

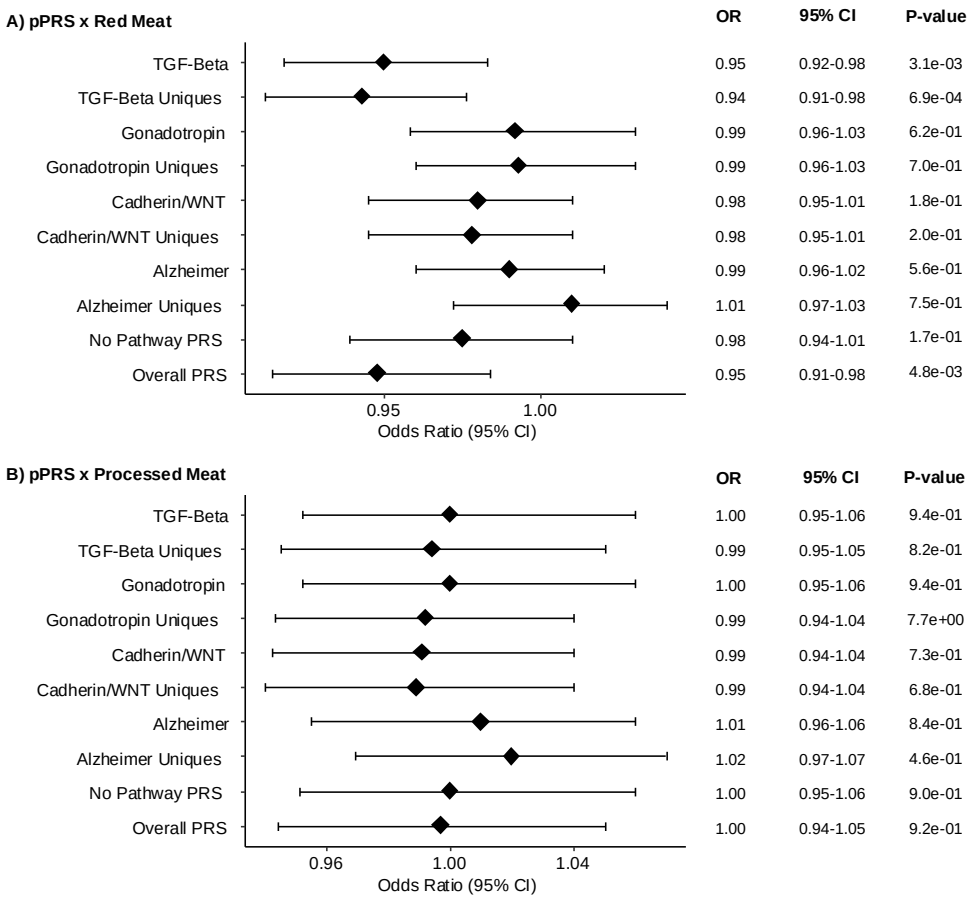

Figure 3. Odds Ratios and 95% CI for pathway-based Polygenic Risk Score interactions (pPRSxE) with **(A)** red meat, and **(B)** processed meat intake in association with colorectal cancer (CRC) risk.
OR; Odds Ratio, CI; Confidence Interval
NOTE: The No Pathway PRS contains variants that have previously shown a genome-wide significant association with colorectal cancer but were not overrepresented in biological pathways. For variants that are exclusively overrepresented in a single pathway, a second subset is generated (i.e., “Uniques”) and the appropriate pPRS computed. P-value corresponds to a 1-degree-of-freedom test of pPRSxE.
All models were adjusted for study, age, sex, and the first three principal components.

Table 1. Association between red meat/processed meat intake with Colorectal Cancer risk stratified by quartiles of pPRS.

	Q1		Q2		Q3		Q4	
Red Meat	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
TGF-β-PRS	1.37 (1.26-1.49)	1.20E-13	1.33 (1.22-1.45)	4.39E-11	1.27 (1.16-1.38)	5.45E-08	1.17 (1.07-1.27)	4.41E-04
TGF-β-$PRS_{uniques}$[a]	1.36 (1.25-1.48)	8.65E-13	1.31 (1.20-1.42)	8.26E-10	1.32 (1.21-1.44)	1.19E-10	1.36 (1.06-1.26)	8.64E-04
Overall PRS	1.32 (1.21-1.44)	1.22E-09	1.32 (1.21-1.44)	1.14E-10	1.35 (1.24-1.47)	2.83E-12	1.17 (1.07-1.28)	5.02E-04
No Pathway PRS[b]	1.29 (1.18-1.41)	3.09E-08	1.3 (1.19-1.41)	1.92E-09	1.33 (1.23-1.45)	1.99E-11	1.24 (1.13-1.35)	1.61E-06
Processed Meat								
TGF-β-PRS	1.39 (1.21-1.6)	3.68E-06	1.5 (1.31-1.73)	9.67E-09	1.3 (1.13-1.5)	1.93E-04	1.42 (1.23-1.63)	9.85E-07
TGF-β-$PRS_{uniques}$[a]	1.47 (1.28-1.69)	5.56E-08	1.32 (1.15-1.52)	1.00E-04	1.48 (1.29-1.7)	2.44E-08	1.31 (1.14-1.51)	1.21E-04
Overall PRS	1.34 (1.16-1.55)	8.40E-05	1.48 (1.28-1.70)	5.43E-08	1.43 (1.24-1.65)	5.04E-07	1.27 (1.10-1.47)	1.09E-03
No Pathway PRS[b]	1.34 (1.16-1.55)	8.15E-05	1.56 (1.35-1.79)	5.63E-10	1.39 (1.21-1.60)	3.68E-06	1.27 (1.10-1.47)	1.05E-03

pPRS, Pathway-based Polygenic Risk Score (PRS); Q1, First Quartile; Q2, Second Quartile; Q3, Third Quartile; Q4, Fourth Quartile; OR, Odds Ratio; CI, Confidence Interval.
[a]pPRS constructed with variants that are only overrepresented in the TGF-β signaling pathway.
[b]PRS constructed with 174 SNPs that did not showed pathway overrepresentation.
Note: All models were adjusted for study, sex, age and the first three Principal Components.

Table 2. Multiplicative interactions between TGF-β-PRS and red meat intake (pPRSxE) in association with colorectal cancer risk.

	TGF-β-PRS x Red Meat		
	OR	95% CI	P-interaction[a]
Model 1	0.950	0.918-0.983	0.0031
Model 2	0.951	0.919-0.984	0.0040
Model 3	0.951	0.919-0.984	0.0041
Model 4	0.949	0.916-0.982	0.0030
Model 5	0.987	0.954-1.021	0.4570

PRS, Polygenic Risk Score; OR, Odds Ratio; CI, Confidence Interval.
Model 1: Adjusted for study, sex, age and the first three Principal Components.
Model 2: Adjusted for covariates in Model 1, plus PRS that contains variants that were not overrepresented in the TGF-β pathway, but that showed statistically significant overrepresentation in other pathways (n = 16)
Model 3: Adjusted for covariates in Model 1, plus PRS that contains SNPs that do not exhibit overrepresentation for the TGF-β pathway but demonstrate overrepresentation in other pathways (n = 25).
Model 4: Adjusted for covariates in Model 1, plus a PRS constructed with all SNPs that do not showcase overrepresentation for the TGF-β pathway but have been marginally associated with the trait (n = 190).
Model 5: pPRS constructed with common variants between TGF-β, and Gonadotropin releasing hormone pathways (n = 9). Adjusted for covariates in Model 1.
[a] Computed with a 1 degree-of-freedom test.

Table 3. Gene-environment interactions (GxE) between single nucleotide polymorphisms (SNP) uniquely overrepresented in the TGF-β pathway[a] and red meat intake in relation to colorectal cancer risk.

SNP	Chr	BP Position	Effect	Gene	Ref	Alt	Alt Allele Freq (1000G EUR)	SNP x Red Meat	
								OR (95% CI)	P-interaction[b]
rs3809570	15	67000117	Intronic/Upstream	*SMAD6*	C	A	0.249	1.013 (0.958-1.071)	0.6580
rs2337113	18	46452327	Intronic	*SMAD7*	A	G	0.450	1.088 (1.037 -1.142)	0.0005
rs2208603	6	55577214	Intergenic	*BMP5*	T	C	0.708	0.946 (0.898-0.996)	0.0360
rs62404966	6	55712124	Intronic	*BMP5*	C	T	0.245	0.962 (0.909-1.019)	0.1830
rs6912214	6	55721302	Intronic	*BMP5*	T	C	0.458	1.019 (0.971-1.070)	0.4430

SNP, single nucleotide polymorphism; Chr, chromosome; BP Position, base pair position based on GRCh37; OR, Odds Ratio; CI, Confidence Interval.

[a]SNPs that do not show overlap with any other pathway, except TGF-β.

[b]Computed with a 1 degree-of-freedom test.